

# ON LEAST FAVORABLE CONFIGURATIONS FOR STEP-UP-DOWN TESTS

Gilles Blanchard, Thorsten Dickhaus, Etienne Roquain and Fanny Villers

*Potsdam University, Humboldt University and UPMC University*

*Abstract:* This paper investigates an open issue related to false discovery rate (FDR) control of step-up-down (SUD) multiple testing procedures. It has been established in earlier literature that for this type of procedure, under some broad conditions, and in an asymptotical sense, the FDR is maximum when the signal strength under the alternative is maximum. In other words, so-called “Dirac uniform configurations” are asymptotically *least favorable* in this setting. It is known that this property also holds in a non-asymptotical sense (for any finite number of hypotheses), for the two extreme versions of SUD procedures, namely step-up and step-down (with extra conditions for the step-down case). It is therefore very natural to conjecture that this non-asymptotical *least favorable configuration* property could more generally be true for all “intermediate” forms of SUD procedures. We prove that this is, somewhat surprisingly, not the case. The argument is based on the exact calculations proposed earlier by Roquain and Villers (2011a), that we extend here by generalizing Steck’s recursion to the case of two populations. Secondly, we quantify the magnitude of this phenomenon by providing a nonasymptotic upper-bound and explicit vanishing rates as a function of the total number of hypotheses.

*Key words and phrases:* False discovery rate, least favorable configuration, multiple testing, Steck’s recursions, step-up-down.

## 1 Introduction

In mathematical statistics, so-called least favorable parameter configurations (LFCs) play a pivotal role. For a given statistical decision problem over a parameter space  $\Theta$  and a given decision rule  $\delta$ , we define an LFC  $\theta^*(\delta)$  as any element of  $\Theta$  that maximizes the risk (expected loss) of  $\delta$  under this parameter, i.e.,

$$\forall \theta \in \Theta, R(\theta, \delta) \leq R(\theta^*(\delta), \delta),$$

where  $R(\theta, \delta)$  denotes the risk of rule  $\delta$  under  $\theta$ . When available, the knowledge of an LFCs allows one to obtain a bound on the risk over a possibly very large parameter space, including non- or semi-parametric cases where  $\Theta$  has infinite dimensionality. Theoretical investigations of minimax properties can rely on the computation of an LFC. Such knowledge is also relevant for practice, because a user of the procedure can be provided with a performance guarantee if an LFC is known. In this case, even if the risk under the LFC cannot be computed in closed form, it can be approximated by a Monte-Carlo method simulating the distribution corresponding to the LFC. Finally, if the parameter space  $\Theta$  is partitioned into disjoint, restricted submodels, it can be of interest to gain knowledge of the LFC of a decision rule  $\delta$  separately over each submodel, thus providing finer-grained information.

LFC considerations naturally occur in hypothesis testing problems. A classical example is that of one-sided tests over a one-dimensional parameter space admitting an isotonic likelihood ratio: it is well-known that the LFC for the type I error probability is located at the boundary of the null hypothesis. This fact is used to derive critical values for uniformly more powerful tests in that setting.

The LFC problem is particularly delicate for multiple hypotheses testing and the latter has been investigated by many authors in previous literature (Finner and Roters, 2001; Benjamini and Yekutieli, 2001; Lehmann and Romano, 2005; Finner, Dickhaus, and Roters, 2007; Romano and Wolf, 2007; Guo and Rao, 2008; Somerville and Hemmelmann, 2008; Finner, Dickhaus, and Roters, 2009; Finner and Gontscharuk, 2009; Gontscharuk, 2010). In that setting, a family of  $m \geq 2$  null hypotheses  $H_1, \dots, H_m$  is to be tested simultaneously under the scope of a common statistical model with parameter space  $\Theta$ , and some type I error criterion is used that accounts for multiplicity. For theoretical as well as practical applications, it is relevant to determine LFCs over the restricted parameter spaces  $\Theta_{m,m_0}$  where exactly  $m_0$  out of  $m$  of the null hypotheses are true. In this setting, LFC results can be derived straightforwardly only in special situations. In the present work, we restrict our attention to multiple testing procedures that depend on the observed data only through a collection of marginal  $p$ -values, each associated to an individual null hypothesis. This is a commonly used setting for multiple testing problems in high dimension. Moreover, we consider procedures

that reject exactly those null hypotheses having their  $p$ -value less than a certain common threshold  $t^*$ , which can possibly be data-dependent. That is,  $t^*$  may depend in a complex way of the entire family of  $p$ -values. We call such procedures threshold-based for short.

In this setting, LFCs crucially depend on the type I error criterion considered. One frequently encountered family of such criteria is given through loss functions that only depend on the number of type I errors, denoted

$$V_m = V_m(\theta, \delta) := |\{1 \leq i \leq m : H_i \text{ is true for } \theta \text{ and gets rejected by } \delta\}|. \quad (1)$$

In other words, the risk takes the form  $R(\theta, \delta) = \mathbb{E}_\theta[\phi(V_m)]$ . Natural assumptions are that  $t^*$  is a nonincreasing function in each  $p$ -value and that  $\phi$  is a non-decreasing function. Then, by additionally assuming that the  $p$ -values are jointly independent, it is known that the LFC over  $\Theta_{m,m_0}$  is a *Dirac-uniform* (DU) distribution (introduced by Finner and Roters, 2001), i. e., such that  $p$ -values corresponding to true nulls are independent uniform variables, while  $p$ -values under alternatives follow a Dirac distribution with point mass 1 in zero. This result is formally restated in Appendix C. For example, this LFC property holds true under the above assumptions for the  $k$ -family-wise error rate ( $k$ -FWER). For a given  $\theta \in \Theta$ , the  $k$ -FWER under  $\theta$  is defined by  $\text{FWER}_{k,\theta} := \mathbb{P}_\theta(V_m \geq k)$ . Strong control of the (1-)family-wise error rate, i. e., ensuring that  $\sup_\theta \text{FWER}_{1,\theta} \leq \alpha$  for a pre-defined level  $\alpha \in (0, 1)$ , is the usual type I error concept in traditional multiple hypotheses testing theory.

However, over the last two decades, progress in application fields such as genomics, proteomics, neuroimaging, and astronomy has lead to massive multiple testing problems with very large systems of hypotheses (Dudoit and van der Laan, 2008; Pantazis, Nichols, Baillet, and Leahy, 2005; Miller, Genovese, Nichol, Wasserman, Connolly, Reichart, Hopkins, Schneider, and Moore, 2001). In this type of applications, ( $k$ -)FWER control is typically too strict a requirement, and a less stringent notion of type I error control is needed in order to ensure reasonable power of corresponding multiple tests. In particular, the false discovery rate (FDR) introduced by Benjamini and Hochberg (1995) has become a standard criterion for type I error control in large-scale multiple testing problems. The FDR is defined as the expected proportion of type I errors among all rejections. Unfortunately, it does not fall into the class of type I error measures considered

in the previous paragraph, so that the above result does not apply. Furthermore, because the average ratio of two dependent random variables is not necessarily increasing in the value of the numerator, the LFC problem for the FDR criterion turns out to be a challenging issue – even for simple classes of multiple tests and under independence assumptions.

In this work, we contribute to the theory of LFCs under the FDR criterion for so-called step-up-down multiple tests (SUD procedures, for short). These procedures constitute an important and wide subclass of threshold-based multiple testing procedures, wherein the threshold  $t^*$  is obtained by comparing the re-ordered  $p$ -values to a fixed set of critical values (see Tamhane, Liu, and Dunnett, 1998; Sarkar, 2002). Furthermore, recent research has reinforced the interest of this type of procedures. For instance, Finner et al. (2009) have shown that step-up-down tests can be used in association with the so-called asymptotically optimal rejection curve (AORC) to provide an asymptotically (as  $m \rightarrow \infty$ ) valid FDR control which is additionally optimal in some specific sense.

Namely, the contributions of the paper are as follows:

- a survey of known LFC results for SUD procedures in specific model classes is provided in Section 3;
- new results and surprising counterexamples for LFCs of SUD procedures are derived in Section 4.
- in Section 5, Steck’s recursion is extended to the case of two populations and we provide a summary of the exact formulas for computing the FDR proposed by Roquain and Villers (2011a,b); these formulas are used to derive the counterexamples previously mentioned.

## 2 Mathematical setting

### 2.1 Models

Given a statistical model, we consider a finite set of  $m \geq 2$  null hypotheses  $H_1, \dots, H_m$ , and a corresponding, fixed collection of tests with associated  $p$ -value family  $\mathbf{p} := (p_i, i \in \{1, \dots, m\})$ . For simplicity, we skip somewhat the formal definition of  $p$ -values and of the underlying statistical model and consider directly

a statistical model for the  $p$ -values, that is, a model for the joint distribution of  $\mathbf{p}$ . In what follows, we denote by  $\mathcal{F}$  the set containing c.d.f.'s from  $[0, 1]$  into  $[0, 1]$  that are continuous.

- The  $p$ -value family  $\mathbf{p}$  follows the *(two group) fixed mixture model* with parameters  $m \geq 2$ ,  $1 \leq m_0 \leq m$  and  $F \in \mathcal{F}$ , for which the corresponding distribution is denoted by  $\text{FM}(m, m_0, F)$ , if  $\mathbf{p} = (p_i, i \in \{1, \dots, m\})$  is a family of mutually independent variables and for all  $i$ ,

$$p_i \sim \begin{cases} U(0, 1) & \text{if } 1 \leq i \leq m_0, \\ F & \text{if } m_0 + 1 \leq i \leq m, \end{cases}$$

where  $U(0, 1)$  denotes the uniform distribution on  $(0, 1)$ .

- The  $p$ -value family  $\mathbf{p}$  follows the *(two group) random mixture model* with parameters  $m \geq 2$ ,  $\pi_0 \in [0, 1]$  and  $F \in \mathcal{F}$ , for which the corresponding distribution is denoted by  $\text{RM}(m, \pi_0, F)$ , if there is an (unobserved) binomial random variable  $m_0 \sim \mathcal{B}(m, \pi_0)$  such that  $\mathbf{p}$  follows the  $\text{FM}(m, m_0, F)$  model conditionally on  $m_0$ . In that case, the  $p$ -values are i.i.d. with (unconditional) c.d.f.  $G(t) = \pi_0 t + (1 - \pi_0)F(t)$ .

In the above definition, note that the true nulls are automatically assigned to the  $m_0$  (random or not) first coordinates. This can be assumed without loss of generality, since we only consider procedures which only depend on  $p$ -values through their reordering in increasing order.

A common additional assumption on  $F$  is that  $F(x) \geq x$  or that  $F$  is concave. For instance, these assumptions are both satisfied in the two following standard examples:

- Gaussian location model:  $F(t) = \bar{\Phi}(\bar{\Phi}^{-1}(t) - \mu)$ , for a given alternative mean  $\mu > 0$ , where  $\bar{\Phi}(z) = \mathbb{P}(Z \geq z)$  for  $Z \sim \mathcal{N}(0, 1)$ . This corresponds to the alternative distribution of  $p$ -values when testing for  $\mu \leq 0$  under a Gaussian location shift model with unit variance.
- Dirac  $\delta_0$  distribution:  $F$  is identically equal to 1, as introduced by Finner and Roters (2001). The corresponding distribution in the FM model is called Dirac-uniform (DU) configuration (or distribution) and denoted by

FM( $m, m_0, F \equiv 1$ ) or simply DU( $m, m_0$ ). We define similarly RM( $m, \pi_0, F \equiv 1$ ). Note that the Dirac-uniform configuration can be seen as an instance of the Gaussian c.d.f. for an alternative mean  $\mu = \infty$ .

In the existing literature, the Dirac-uniform distribution has often be considered as the first candidate for being an LFC of several global type I error rates (with or without a theoretical support) (see, e.g., Finner et al., 2007; Romano and Wolf, 2007; Somerville and Hemmelmann, 2008).

## 2.2 Procedures

In this paper, we consider the particular class of multiple testing procedures called step-up-down procedures, introduced by Tamhane et al. (1998), see also Sarkar (2002). First define a *threshold* or *critical value* collection as any nondecreasing sequence  $\mathbf{t} = (t_k)_{1 \leq k \leq m} \in [0, 1]^m$  (with  $t_0 = 0$  by convention).

**Definition 2.1.** *Let us order the  $p$ -values  $p_{(1)} \leq p_{(2)} \leq \dots \leq p_{(m)}$  (with the convention  $p_{(0)} = 0$ ). For any threshold collection  $\mathbf{t}$ , the step-up-down (SUD) procedure with threshold collection  $\mathbf{t}$  and of order  $\lambda \in \{1, \dots, m\}$ , denoted here by  $SUD_\lambda(\mathbf{t})$ , rejects the  $i$ -th hypothesis if  $p_i \leq t_{\hat{k}}$ , with*

$$\hat{k} = \begin{cases} \max\{k \in \{\lambda, \dots, m\} : \forall k' \in \{\lambda, \dots, k\}, p_{(k')} \leq t_{k'}\} & \text{if } p_{(\lambda)} \leq t_\lambda; \\ \max\{k \in \{0, \dots, \lambda\} : p_{(k)} \leq t_k\} & \text{if } p_{(\lambda)} > t_\lambda. \end{cases} \quad (2)$$

In the sequel, for convenience, we identify procedures with their rejection sets, e.g.,  $SUD_\lambda(\mathbf{t}) = \{1 \leq i \leq m : p_i \leq t_{\hat{k}}\}$ . An important remark is that the cases  $\lambda = 1$  and  $\lambda = m$  correspond to the traditional step-down and step-up procedures, respectively. An illustration is provided in Figure 1.

A classical choice for the threshold collection  $\mathbf{t}$  consists of Simes' (1986) critical values  $t_k = \alpha k/m$  for a pre-specified level  $\alpha \in (0, 1)$ . The corresponding step-up-down procedure is called the *linear step-up-down* procedure and is denoted by LSUD $_\lambda$ . In particular, for  $\lambda = 1$  and  $\lambda = m$ , the procedure LSUD $_\lambda$  is simply denoted by LSD and LSU, respectively. LSU corresponds to the famous linear step-up procedure of Benjamini and Hochberg (1995).

It is common to consider threshold collections of the form  $t_k = \rho(k/m)$  for a function  $\rho : [0, 1] \rightarrow [0, 1]$ . This function is generally assumed to satisfy the

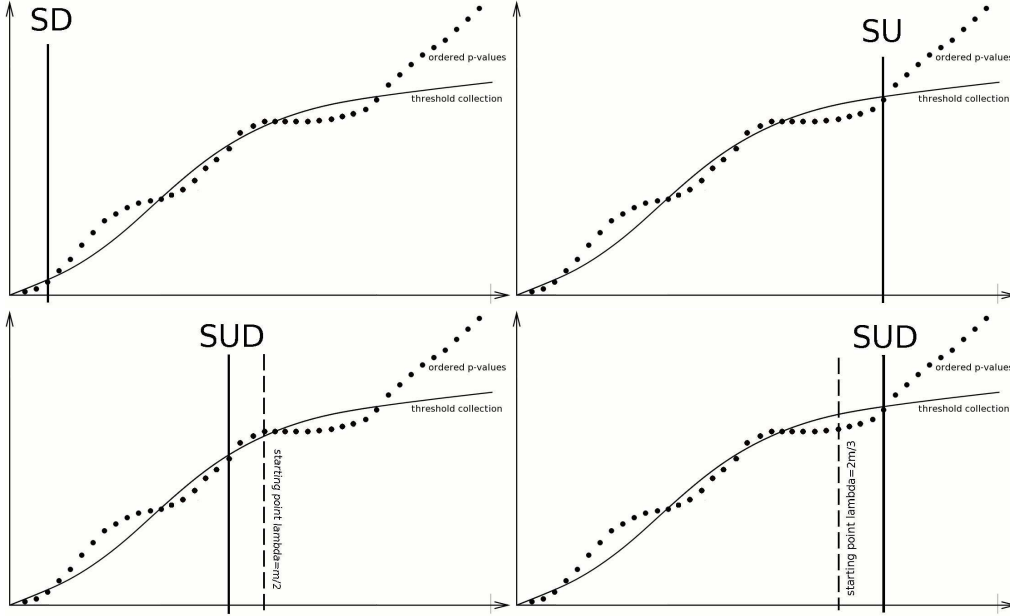


Figure 1: Value of  $\hat{k}$  (vertical solid line), defined by (2), for several procedures of the SUD type. The bottom-right SUD procedure (using  $\lambda = 2m/3$ ) coincides with the SU procedure for this realization of the  $p$ -value family.

following assumptions:

$$\rho : [0, 1] \rightarrow [0, 1] \text{ is continuous and non-decreasing;} \quad (3)$$

$$x \in (0, 1] \rightarrow \rho(x)/x \text{ is non-decreasing.} \quad (4)$$

The function  $\rho$  is called the critical value function (while its inverse is generally called the rejection curve, see e.g. Finner et al., 2009). Observe that assumption (3) can always be assumed to hold when  $m$  is fixed, however it is of interest in the case of an asymptotical analysis as  $m \rightarrow \infty$  (in which case  $\rho$  is assumed to be independent of  $m$ ). Assumption (4) on the other hand restricts the possible threshold collection also for any fixed  $m$ . It will be often used in this paper. For a fixed finite  $m$ , assumptions (3) and (4) taken together are equivalent to “ $k \mapsto t_k/k$  is non-decreasing”.

### 2.3 False discovery rate and LFCs

As introduced by Benjamini and Hochberg (1995), the false discovery rate of a multiple testing procedure is defined as the averaged ratio of the number of erroneous rejections to the total number of rejections. In our setting, for a distribution  $P$  being either  $\text{FM}(m, m_0, F)$  or  $\text{RM}(m, \pi_0, F)$ , the FDR of a step-up-down procedure can be written as

$$\text{FDR}(\text{SUD}_\lambda(\mathbf{t}), P) = \mathbb{E}_{\mathbf{p} \sim P} [\text{FDP}(\text{SUD}_\lambda(\mathbf{t}), m_0, \mathbf{p})], \quad (5)$$

for which the FDP is the false discovery proportion defined by

$$\text{FDP}(\text{SUD}_\lambda(\mathbf{t}), m_0, \mathbf{p}) = \frac{|\{1 \leq i \leq m_0 : p_i \leq t_k\}|}{|\{1 \leq i \leq m : p_i \leq t_k\}| \vee 1}, \quad (6)$$

where  $|\cdot|$  denotes the cardinality function and in which  $m_0$  is either fixed or random whether  $P$  is  $\text{FM}(m, m_0, F)$  or  $\text{RM}(m, \pi_0, F)$ , respectively. For short, the quantity  $\text{FDR}(\text{SUD}_\lambda(\mathbf{t}), \text{FM}(m, m_0, F))$  is often denoted  $\text{FDR}(\text{SUD}_\lambda(\mathbf{t}), m_0, F)$ , or  $\text{FDR}(\text{SUD}_\lambda(\mathbf{t}), F)$  when the context makes the interpretation unambiguous. Similarly,  $\text{FDR}(\text{SUD}_\lambda(\mathbf{t}), \text{RM}(m, \pi_0, F))$  can be shortened as  $\text{FDR}(\text{SUD}_\lambda(\mathbf{t}), \pi_0, F)$ , or  $\text{FDR}(\text{SUD}_\lambda(\mathbf{t}), F)$ .

**Definition 2.2.** Any  $F' \in \mathcal{F}$  is called a *least favorable configuration (LFC)* for the FDR of  $\text{SUD}_\lambda(\mathbf{t})$  in a fixed mixture model with  $m_0$  true hypotheses out of  $m$  and over the class  $\mathcal{F}$  if

$$\forall F \in \mathcal{F}, \quad \text{FDR}(\text{SUD}_\lambda(\mathbf{t}), m_0, F) \leq \text{FDR}(\text{SUD}_\lambda(\mathbf{t}), m_0, F').$$

A similar definition holds for a random mixture model with  $m$  hypotheses and proportion  $\pi_0$  of true hypotheses.

The above definition can possibly be restricted to a subclass  $\mathcal{F}' \subset \mathcal{F}$  (typically, the class of concave c.d.f.s). This will be clearly specified in the context.

Obviously, if  $F'$  is an LFC for the fixed mixture model for all values of  $m_0$ , then it is also an LFC in the  $\text{RM}(m, \pi_0, F)$  model for any value of  $\pi_0$  (by integrating over  $m_0 \sim \mathcal{B}(m, \pi_0)$ ).



### 3 Survey of known LFCs under the FDR criterion

Recent results about LFCs for the FDR criterion related to step-up-down type procedures are summarized in Figure 2 and explained below. These results hold either under the fixed mixture model or the random mixture model, hence involve a maximization over distributions where the  $p$ -values are independent. (While the present paper is focused on this setting, let us mention here briefly, that LFCs for the FDR criterion under *arbitrary* dependencies have also been studied, see, e.g. Lehmann and Romano, 2005; Guo and Rao, 2008.)

First, let us consider the problem of the monotonicity of  $\text{FDR}(\text{SUD}_\lambda(\mathbf{t}))$  in  $\lambda$  (vertical arrows). Recently, it was proved that, whenever  $F$  is concave, the FDR grows as the rejection set grows (Zeisel, Zuk, and Domany, 2011, Theorem 4.1). Interestingly, the rejection set  $R$  can have a very general form: the only condition is that  $|R|$  is a measurable function of the order statistics of the family of  $p$ -values under consideration. From (2) and since for any  $\lambda \in \{1, \dots, m-1\}$ , the rejection set of  $\text{SUD}_\lambda(\mathbf{t})$  is included in the one of  $\text{SUD}_{\lambda+1}(\mathbf{t})$ , we obtain that for a concave  $F$ ,

$$\text{FDR}(\text{SUD}_\lambda(\mathbf{t})) \leq \text{FDR}(\text{SUD}_{\lambda+1}(\mathbf{t})),$$

both for  $\text{FM}(m, m_0, F)$  and  $\text{RM}(m, \pi_0, F)$  models. This implies in particular that  $\text{FDR}(\text{SD}(\mathbf{t})) \leq \text{FDR}(\text{SU}(\mathbf{t}))$  for a concave  $F$ . Other studies establish similar inequalities, but with a condition on the threshold collection  $\mathbf{t}$ , not on  $F$ . Precisely, Theorem 4.3 of Finner et al. (2009) and Theorem 3.10 of Gontscharuk (2010) establish that, when  $k \mapsto t_k/k$  is nondecreasing, for any  $\lambda \in \{1, \dots, m-1\}$ ,

$$\text{FDR}(\text{SUD}_\lambda(\mathbf{t})) \leq \text{FDR}(\text{SU}(\mathbf{t})),$$

both for  $\text{FM}(m, m_0, F)$  and  $\text{RM}(m, \pi_0, F)$  models. In particular, the fact that  $\text{FDR}(\text{SU}(\mathbf{t}))$  dominates the FDR of  $\text{SD}(\mathbf{t})$  is quite well established in multiple testing literature. Nevertheless, let us stress that this is no longer the case for “atypical” configurations of  $F$  and  $\mathbf{t}$ , as we state in Appendix B.

Secondly, let us consider the monotonicity of  $\text{FDR}(\text{SUD}_\lambda(\mathbf{t}), F)$  in  $F$ . In the step-up case (i.e.,  $\lambda = m$ ), the situation is somewhat simple: Theorem 5.3 of Benjamini and Yekutieli (2001) states that  $F \leq F'$  implies  $\text{FDR}(\text{SUD}_\lambda(\mathbf{t}), F) \leq \text{FDR}(\text{SUD}_\lambda(\mathbf{t}), F')$  whenever  $k \mapsto t_k/k$  is nondecreasing. Moreover, the inequality is reversed whenever  $k \mapsto t_k/k$  is nonincreasing. In the step-down case (i.e.,

$\lambda = 1$ ) and for a  $\text{RM}(m, \pi_0, F)$  model, Theorem 4.1 of Roquain and Villers (2011a) states that the Dirac-uniform configuration ( $F \equiv 1$ ) is an LFC under some complex condition on the threshold collection  $\mathbf{t}$ , that is fulfilled by the linear threshold collection  $t_k = \alpha k/m$ ,  $\alpha \in (0, 1)$  and over the class of concave c.d.f.'s. However, for  $\lambda \notin \{1, m\}$  (i.e., an “intermediate” SUD procedure), finding LFC's is more delicate and the only known result is asymptotic, as  $m$  tends to infinity. Precisely, combining Theorem 4.3 of Finner et al. (2009) and Lemma 3.7 of Gontscharuk (2010), we easily derive the following result:

**Theorem 3.1.** *[Gontscharuk (2010)] Consider a step-up-down procedure using a threshold collection of the form  $t_k = \rho(k/m)$ , where  $\rho$  satisfies (3) and (4). Assume that the step-up-down procedure is performed at an order  $\lambda = \lambda_m$  such that  $\lambda_m/m \rightarrow \kappa \in [0, 1]$ . Assume that  $m_0/m \rightarrow \zeta \in [0, 1]$  and that, under the  $\text{DU}(m, m_0)$  distribution, the number of rejections of  $\text{SUD}_\lambda(\mathbf{t})$  satisfies that  $|\text{SUD}_\lambda(\mathbf{t})|/m$  converges in probability as  $m$  grows to infinity. Then, in the fixed mixture model  $\text{FM}(m, m_0, F)$ , we have for any  $F \in \mathcal{F}$ ,*

$$\limsup_m \{FDR(\text{SUD}_{\lambda_m}(\mathbf{t}), F) - FDR(\text{SUD}_{\lambda_m}(\mathbf{t}), F \equiv 1)\} \leq 0, \quad (7)$$

*either for all  $\zeta \in [0, 1]$  if  $\kappa > 0$  or for all  $\zeta \in [0, 1]$  if  $\kappa = 0$ .*

However, for a finite  $m$ , and  $\lambda \notin \{1, m\}$  no result is known about LFC's to our knowledge. This is the point of the paper and is symbolized by the question mark in the middle of Figure 2.

Finally, let us consider the linear SUD procedure, that is, the SUD procedure using the threshold collection  $t_k = \alpha k/m$ ,  $\alpha \in (0, 1)$  (corresponding to  $\rho(x) = \alpha x$ ). Since both LSU and LSD procedures satisfy that DU is an LFC and since an SUD procedure can be expressed as a combination of an SU and an SD, we might make the following conjecture, which is the starting point of this paper.

**Conjecture 3.2.** *For any  $m \geq 2$ , the Dirac-uniform configuration ( $F \equiv 1$ ) is a least favorable configuration for the FDR of the linear step-up-down procedure, in the  $\text{RM}(m, \pi_0, F)$  and  $\text{FM}(m, m_0, F)$  models.*

Obviously, a similar conjecture might be formulated for a (non-linear) step-up-down procedure using  $\rho$  satisfying (3) and (4).

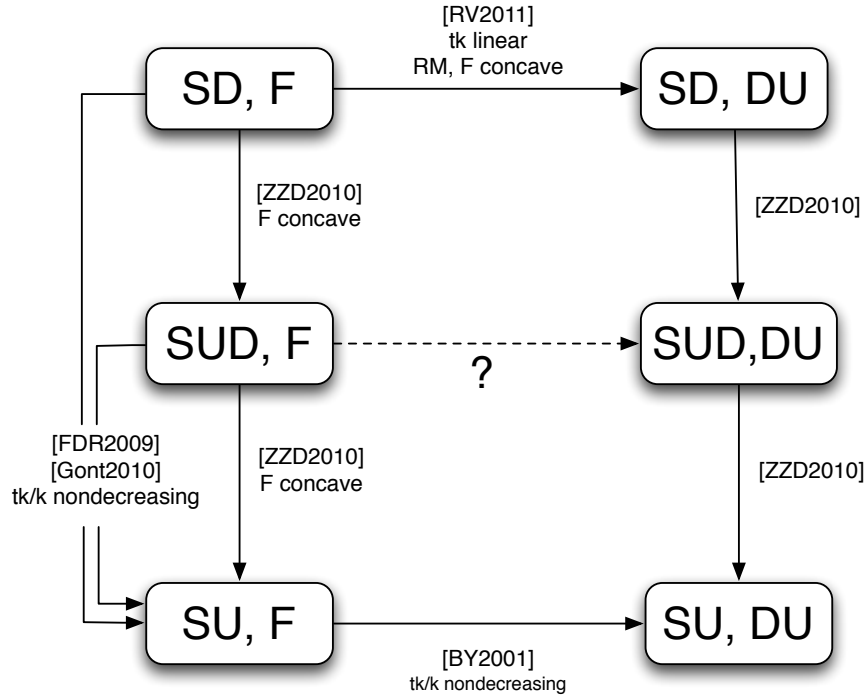


Figure 2: An arrow “ $A \rightarrow B$ ” means “ $\text{FDR}(A) \leq \text{FDR}(B)$ ”. These results hold for the fixed mixture (FM) model except when “RM” is written. The brackets are a shortened reference to the corresponding literature, see main text for more details.

## 4 Investigating Conjecture 3.2

### 4.1 Disproving the conjecture

The exact calculations described in Section 5 allow to compute the value of  $FDR(LSUD_\lambda(\mathbf{t}))$  exactly. This shows the following (numerical) result.

**Proposition 4.1.** *For  $m = 10$ , consider the linear step-up-down procedure  $LSUD_\lambda$  at level  $\alpha = 0.5$  and for  $\lambda \in \{4, 5, 6, 7\}$ . Then, we have*

$$FDR(LSUD_\lambda, F) > FDR(LSUD_\lambda, F \equiv 1), \quad (8)$$

*in either of the two following cases:*

- *in the  $FM(m, m_0, F)$  model, with  $m_0 = 7$  and  $F(x) = x$ ;*
- *in the  $RM(m, \pi_0, F)$  model, with  $\pi_0 = 7/10$  and  $F(x) = x$ .*

This disproves Conjecture 3.2 and shows that finding the LFC for SUD is more difficult than for SU and SD separately. More generally, Figure 3 reports some obtained values in the Gaussian case  $F = \overline{\Phi}(\overline{\Phi}^{-1}(\cdot) - \mu)$  for different values of the alternative mean  $\mu$ . We observe that the result for  $FM(m, m_0, F)$  (left panel) or  $RM(m, \pi_0, F)$  (right panel) are qualitatively the same: there is a range of values for  $\lambda \in \{2, \dots, m-1\}$  for which the FDR is larger for a smaller  $\mu$ . However, this phenomenon seems to decay when  $m$  becomes larger, see Figure 3 for  $m = 100$ . Also, when  $\alpha$  decreases, the phenomenon still occurs but its amplitude decays. To alleviate the concern that this somewhat unexpected phenomenon could be due to numerical inaccuracies in the computation of the exact formulas (which involve several nested recursions), the reported results were double-checked via extensive and independent Monte-Carlo simulations, which confirmed the validity of the reported curves.

### 4.2 Nonasymptotic bound

In the present section, we investigate the amplitude of the phenomenon observed in the previous section as a function of the number of hypotheses. In other words, we derive a more explicit and non-asymptotic version of the limit appearing in (7). For this, we consider a perturbation analysis of the SUD procedure as defined

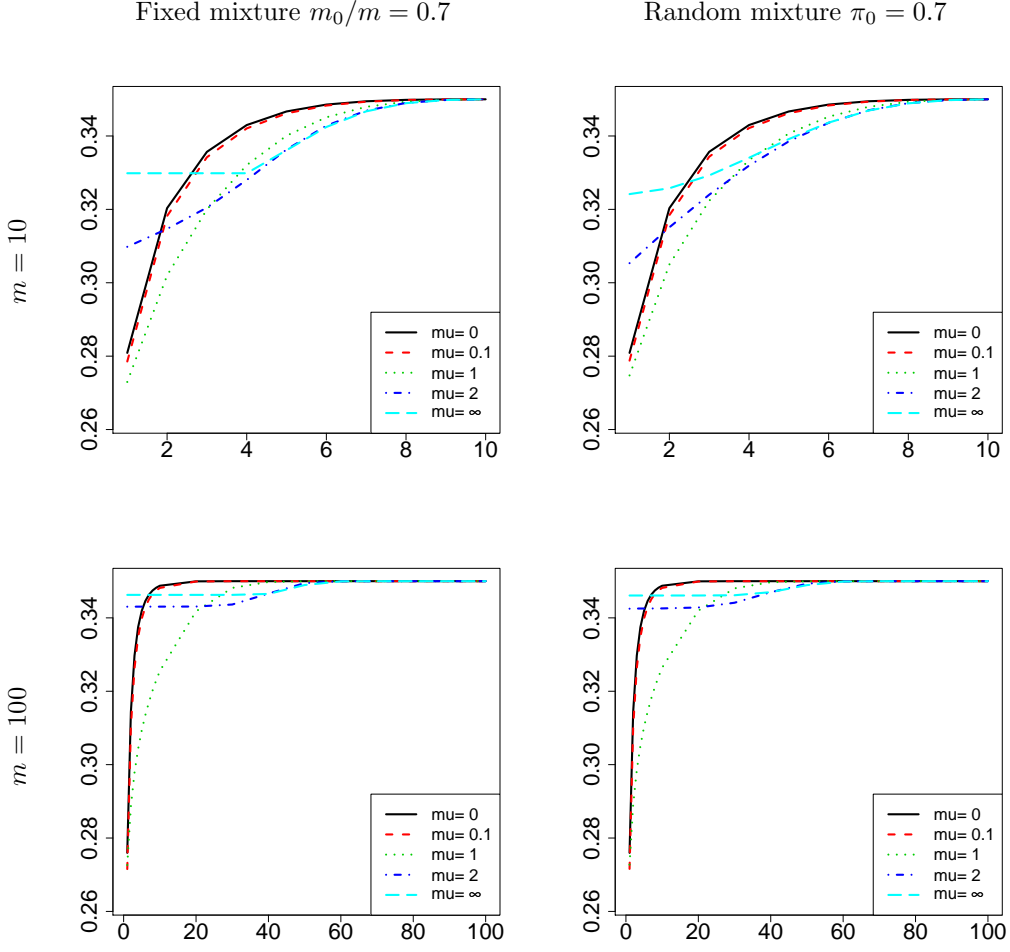


Figure 3: The LFC of LSUD is not always DU.  $\text{FDR}(\text{LSUD}_\lambda)$  as a function of the order  $\lambda \in \{1, \dots, m\}$ . Left: fixed mixture; Right: random mixture. One sided Gaussian location model with parameter  $\mu$ .  $\alpha = 0.5$ .

in (2), under the Dirac-uniform model, and when the empirical c.d.f. of the  $p$ -values is  $\delta$ -close to the population c.d.f. (which happens with large probability). In order to state the result in a compact form, we first introduce the following notation for an SUD threshold in a continuous setting.

**Definition 4.2.** Let  $\rho : [0, 1] \rightarrow [0, 1]$  satisfying (3). For any non-decreasing

function  $G : [0, 1] \rightarrow [0, 1]$ , and  $\ell \in [0, 1]$ , we define

$$\mathcal{U}(\ell, G) = \begin{cases} \min \{u \in [\ell, 1] : G(\rho(u)) \leq u\} & \text{if } G(\rho(\ell)) \geq \ell; \\ \max \{u \in [0, \ell] : G(\rho(u)) \geq u\} & \text{if } G(\rho(\ell)) < \ell. \end{cases} \quad (9)$$

Observe that in the above definition, the infimum and supremum are well-defined since the considered sets are non-empty; using that  $G$  is non-decreasing, it can be seen that  $\mathcal{U}(\ell, G)$  is a fixed point of the function  $G \circ \rho$  (so that the infimum is indeed a minimum and the supremum, a maximum). Unfortunately, the number of rejections  $\hat{k}$  of the SUD procedure as defined in (2) does not always satisfy  $\hat{k}/m = \mathcal{U}(\lambda/m, \hat{\mathbb{G}}_m)$  (because of the step-down part, see Figure 6 in Section 7.2). Nevertheless, the following lemma is proved in Section 7.2.

**Lemma 4.3.** *With the above notation, if the threshold collection  $\mathbf{t}$  is defined as  $t_k = \rho(k/m)$ , we have*

$$\mathcal{U}(\lambda/m, \hat{\mathbb{G}}_m) \leq \hat{k}/m \leq \mathcal{U}(\lambda/m, (\hat{\mathbb{G}}_m + m^{-1}) \wedge 1), \quad (10)$$

where  $\hat{k}$  is defined by (2) and  $\hat{\mathbb{G}}_m(x) := m^{-1} \sum_{i=1}^m \mathbf{1}\{p_i \leq x\}$  is the empirical c.d.f. of the  $p$ -values.

We now state our main result.

**Theorem 4.4.** *Consider a threshold collection  $\mathbf{t}$  of the form  $t_k = \rho(k/m)$ , where  $\rho : [0, 1] \rightarrow [0, 1]$  satisfies (3) and (4). Let  $\zeta \in (0, 1)$  be an arbitrary constant. For  $\delta \in (0, 1)$ , define*

$$u_\delta^+ = \mathcal{U}(\lambda, (G_\zeta^{DU} + \delta) \wedge 1) \quad \text{and} \quad u_\delta^- = \mathcal{U}(\lambda, (G_\zeta^{DU} - \delta) \vee 0),$$

where  $G_\zeta^{DU}(x) := (1 - \zeta) + \zeta x$  (see Figure 4 for an illustration). Let us define the remaining term: for any  $y \in (0, 1)$ ,

$$\varepsilon(\delta, m, \zeta, y) := \frac{\rho(u_\delta^+) - \rho(u_\delta^-)}{u_\delta^+} + \frac{4}{1 - \zeta} e^{-2m(\delta - y - \frac{1}{m})_+^2 (1 - y/\zeta)_+}. \quad (11)$$

Then, for any  $F \in \mathcal{F}$  and  $\lambda \in \{1, \dots, m\}$  the following holds.

- In the  $\text{FM}(m, m_0, F)$  model with  $0 < m_0 < m$  and  $\nu = \max_{k \in \{m_0 - 1, m_0\}} \{|k/m - \zeta|\} \in [0, 1]$ , we have

$$\text{FDR}(\text{SUD}_\lambda(\mathbf{t}), m_0, F) \leq \text{FDR}(\text{SUD}_\lambda(\mathbf{t}), m_0, F \equiv 1) + \frac{m_0}{m} \varepsilon(\delta, m, \zeta, \nu); \quad (12)$$

- In the  $\text{RM}(m, \pi_0, F)$  model with  $\pi_0 = \zeta$ , we have for any  $\gamma \in (0, 1)$ ,

$$FDR(SUD_\lambda(\mathbf{t}), \pi_0, F) \leq FDR(SUD_\lambda(\mathbf{t}), \pi_0, F \equiv 1) + \pi_0 \varepsilon(\delta, m, \zeta, \gamma) + 4e^{-2m(\gamma-1/m)_+^2}. \quad (13)$$

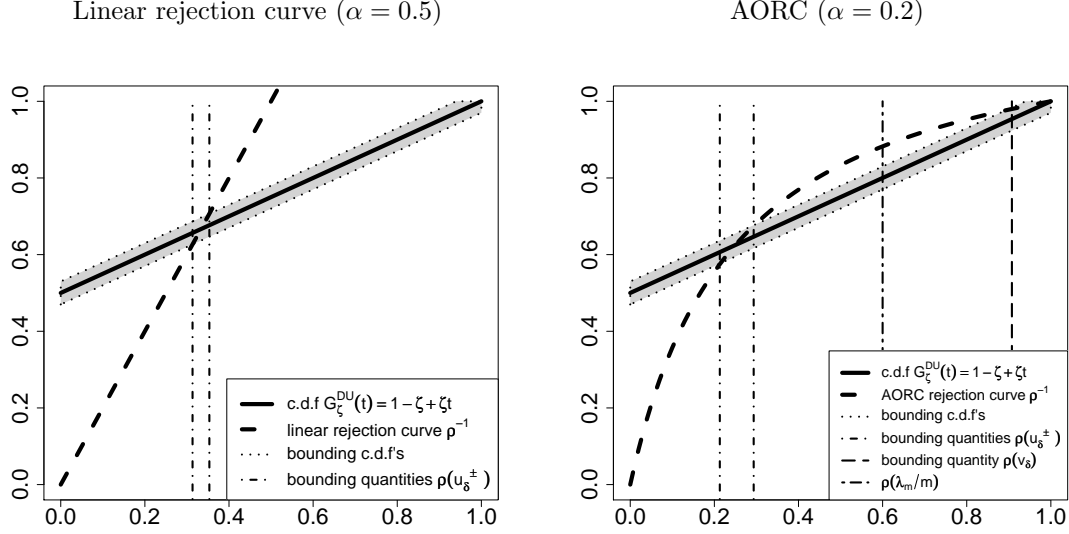


Figure 4: Illustration for  $u_\delta^+$  and  $u_\delta^-$  in the case of the LSUD and the SUD based on AORC. Here the X-axis is on the “threshold scale”  $t = \rho(u)$ .  $\zeta = 0.5$ ;  $\delta = 0.03$ . The area between  $(G_\zeta^{DU} - \delta) \vee 0$  and  $(G_\zeta^{DU} + \delta) \wedge 1$  is displayed in gray.

Theorem 4.4 is proved in Section 7.3. This result is illustrated in the two following examples.

1. Let us first apply Theorem 4.4 in the particular case where  $\rho(x) = \alpha x$  is the linear critical value function, see the left panel of Figure 4. In that case, we have  $u_\delta^- = \frac{1-\zeta-\delta}{1-\alpha\zeta} \vee 0$  and  $u_\delta^+ = \frac{1-\zeta+\delta}{1-\alpha\zeta} \wedge 1$ . Hence,

$$(\rho(u_\delta^+) - \rho(u_\delta^-))/u_\delta^+ \leq \frac{2\alpha\delta}{1-\zeta+\delta}.$$

As a result, (12) and (13) hold by replacing  $(\rho(u_\delta^+) - \rho(u_\delta^-))/u_\delta^+$  by  $\frac{2\alpha\delta}{1-\zeta+\delta}$  inside the remaining term  $\varepsilon(\delta, m, \zeta, y)$ .

2. Second, for  $\zeta > \alpha$ , let us use

$$\rho(u) = \alpha u / (1 - u(1 - \alpha)), \text{ that is, } \rho^{-1}(t) = t / (\alpha + t(1 - \alpha)). \quad (14)$$

The rejection curve  $\rho^{-1}$ , displayed in the right panel of Figure 4, is called the “asymptotically optimal rejection curve” (AORC). It was introduced by Finner et al. (2009) for the purpose of improving the power of the linear critical value function. When applying Theorem 4.4 with this choice of  $\rho$ , the calculation of  $u_\delta^-$  and  $u_\delta^+$  depends on the position of the parameter  $\lambda_m/m$  on  $[0, 1]$  and  $\rho(u_\delta^+) - \rho(u_\delta^-)$  may not vanish when  $\delta$  becomes small, see Figure 4. Fortunately, when  $\rho(\lambda_m/m)$  (dotted-long-dashed line) is smaller than the quantity  $\rho(v_\delta)$  (dashed line), the two points  $\rho(u_\delta^-)$  and  $\rho(u_\delta^+)$  (dotted-dashed lines) are expected to be close as  $\delta$  becomes small and the bound given in Theorem 4.4 will vanish. The exact expressions of  $u_\delta^-$ ,  $u_\delta^+$  and  $v_\delta$  can be easily derived by solving the corresponding quadratic equations. For short, we only report the equivalent as  $\delta$  tends to 0:

$$\begin{aligned} v_\delta &= 1 - \delta\alpha/(\zeta - \alpha) + O(\delta^2) \\ u_\delta^+ &= (1 - \zeta)/(1 - \alpha) + \delta\zeta/(\zeta - \alpha) + O(\delta^2) \\ u_\delta^- &= (1 - \zeta)/(1 - \alpha) - \delta\zeta/(\zeta - \alpha) + O(\delta^2). \end{aligned}$$

Since  $\rho'((1 - \zeta)/(1 - \alpha)) = \alpha/\zeta^2$ , we have  $\rho(u_\delta^+) - \rho(u_\delta^-) = 2\alpha\delta/(\zeta^2 - \alpha\zeta) + O(\delta^2)$ . As a result, assuming  $\zeta > \alpha$  and  $\lambda/m < v_\delta$ , we can derive that (12) and (13) hold and that quantity  $(\rho(u_\delta^+) - \rho(u_\delta^-))/u_\delta^+$  is equivalent to  $\frac{2\alpha(1-\alpha)}{(\zeta^2-\alpha\zeta)(1-\zeta)}\delta$  (as  $\delta$  tends to zero) in the remaining term  $\varepsilon(\delta, m, \zeta, y)$ .

### 4.3 Convergence rate when $m \rightarrow \infty$

We can now use Theorem 4.4 in an asymptotic sense and for specific critical value functions, in order to obtain an explicit bound on the convergence rate of the limit appearing in (7).

**Corollary 4.5.** *Let  $\alpha \in (0, 1)$ . Consider a threshold collection of the form  $t_k = \rho(k/m)$ , where  $\rho$  is either  $\rho(x) = \alpha x$  (linear) or given by (14), that is, associated with the AORC. Consider the SUD procedure of threshold collection  $\mathbf{t}$  and of order  $\lambda = \lambda_m$  possibly depending on  $m$ . Consider either the  $\text{FM}(m, m_0, F)$  model with*



$m_0 = \lfloor \zeta_m m \rfloor$  or  $\text{RM}(m, \pi_0, F)$  model with  $\pi_0 = \zeta_m$ , for some  $\zeta_m \in (\alpha, 1)$  possibly depending on  $m$ . Assume that  $\frac{1}{1-\zeta_m} \sqrt{(\log m)/m} = o(1)$ . For the AORC case, assume moreover  $\frac{1}{1-\lambda_m/m} \sqrt{(\log m)/m} = o(1)$ . Then we have for any  $F \in \mathcal{F}$ ,

$$(FDR(SUD_{\lambda_m}(\mathbf{t}), F) - FDR(SUD_{\lambda_m}(\mathbf{t}), F \equiv 1))_+ = O\left(\frac{1}{1-\zeta_m} \sqrt{\frac{\log m}{m}}\right). \quad (15)$$

Corollary 4.5 is proved in Section 7.4 and is an easy consequence of Theorem 4.4 when taking  $\delta = \delta_m$  (and  $\gamma = \gamma_m$ ) suitably tending to zero. Assuming  $\zeta_m > \alpha$  is not an important restriction because when  $\zeta_m \leq \alpha$ , controlling the FDR is a trivial problem: the procedure rejecting all the hypotheses has an FDR (and even an FDP) smaller than  $\zeta_m \leq \alpha$ .

While focusing on the linear and AORC rejection curve, the conclusion of Corollary 4.5 is substantially more informative than Theorem 3.1: it evaluates what is the order of the error when considering that the DU is an LFC of an SUD test. For  $\zeta_m = \zeta \in (0, 1)$  fixed with  $m$ , note that the rate of convergence in (15) is equal to the parametric rate, up to a  $\log m$  factor. Furthermore, the constant in the  $O(\cdot)$  can be derived explicitly by using the bound from the previous section. For  $\zeta_m$  tending to 1 (not too quickly, “fairly” sparse case), the convergence rate is slower.

As a counterpart, assumptions of Corollary 4.5 are more restrictive than those of Theorem 3.1. In particular, they exclude the case where  $\zeta_m$  tends to 1 faster than  $\sqrt{(\log m)/m}$  (“highly” sparse case). This is a limitation of the methodology used to prove the nonasymptotic results. This problem may possibly be fixed by adapting the proof of Lemma 3.7 of Gontscharuk (2010) to a nonasymptotic setting, but this falls outside of the intended scope of this paper.

## 5 Exact formulas

In this section, we gather some of the formulas derived by Roquain and Villers (2011a,b) to calculate the joint distribution of the number of false discoveries and the number of discoveries. Moreover, we complement this work by giving a new recursion that makes these formulas fully usable. These calculations are used to state Proposition 4.1.

## 5.1 A new Steck's recursion

For any  $k \geq 0$  and any threshold collection  $\mathbf{t} = (t_1, \dots, t_k)$ , we denote

$$\Psi_k(\mathbf{t}) = \Psi_k(t_1, \dots, t_k) = \mathbb{P} [U_{(1)} \leq t_1, \dots, U_{(k)} \leq t_k]. \quad (16)$$

where  $(U_i)_{1 \leq i \leq k}$  is a sequence of  $k$  variables i.i.d. uniform on  $(0, 1)$  and with the convention  $\Psi_0(\cdot) = 1$ . In practice, quantity (16) can be evaluated using standard Steck's recursion  $\Psi_k(\mathbf{t}) = (t_k)^k - \sum_{j=0}^{k-2} \binom{k}{j} (t_k - t_{j+1})^{k-j} \Psi_j(t_1, \dots, t_j)$  (Shorack and Wellner, 1986, p. 366–369).

Next, we generalize the latter to the case of two populations. Define for  $0 \leq k_0 \leq k$  and any threshold collection  $\mathbf{t} = (t_1, \dots, t_k)$ ,

$$\Psi_{k,k_0,F}(t_1, \dots, t_k) = \mathbb{P} [U_{(1)} \leq t_1, \dots, U_{(k)} \leq t_k], \quad (17)$$

where  $(U_i)_{1 \leq i \leq k}$  is a sequence of  $k$  variables such that  $(U_i)_{1 \leq i \leq k_0}$  are i.i.d. uniform on  $(0, 1)$ , independently of  $(U_i)_{k_0+1 \leq i \leq k}$  i.i.d. of c.d.f.  $F$  and with the convention  $\Psi_{0,0,F}(\cdot) = 1$ . The computation of  $\Psi_{k,k_0,F}$  is more difficult than  $\Psi_k$  because it involves non i.i.d. variables. To our knowledge the existing formulas for computing  $\Psi_{k,k_0,F}$  have a complexity exponential with  $k$  (Glueck, Karimpour-Fard, Mandel, Hunter, and Muller, 2008). Here, we propose a substantially less complex computation, by generalizing Steck's recursions as follows.

**Proposition 5.1.** *The following recursion holds: for  $0 \leq k_0 \leq k$ ,*

$$\begin{aligned} \Psi_{k,k_0,F}(t_1, \dots, t_k) = & (t_k)^{k_0} F(t_k)^{k-k_0} - \sum_{\substack{0 \leq j_0 \leq j \leq k-2 \\ j_0 \leq k_0 \\ j-j_0 \leq k-k_0}} \binom{k_0}{j_0} \binom{k-k_0}{j-j_0} \\ & \times (t_k - t_{j+1})^{k_0-j_0} (F(t_k) - F(t_{j+1}))^{k-k_0-j+j_0} \Psi_{j,j_0,F}(t_1, \dots, t_j), \end{aligned} \quad (18)$$

with the convention  $0^0 = 1$ .

This is proved in Section 7.1. Note that the case  $k = k_0$  reduces to the standard (one population) Steck's recursion.

## 5.2 FDR formulas

Using the  $\Psi_k$ 's and  $\Psi_{k,k_0}$ 's, let us define the following useful quantities: for any threshold collection  $\mathbf{t} = (t_k)_{1 \leq k \leq m}$ ,  $F \in \mathcal{F}$ ; for any  $\pi_0 \in [0, 1]$ ,  $k \geq 0$ ,  $k \leq m$ ,

$0 \leq j \leq k$ , we let

$$\begin{aligned} \mathcal{P}_{m,\pi_0,F}(\mathbf{t}, k, j) &= \binom{m}{j} \binom{m-j}{k-j} \pi_0^j \pi_1^{k-j} (t_k)^j (F(t_k))^{k-j} \\ &\quad \times \Psi_{m-k}(1 - G(t_m), \dots, 1 - G(t_{k+1})); \end{aligned} \quad (19)$$

$$\begin{aligned} \tilde{\mathcal{P}}_{m,\pi_0,F}(\mathbf{t}, k, j) &= \binom{m}{j} \binom{m-j}{k-j} \pi_0^j \pi_1^{k-j} (1 - G(t_{k+1}))^{m-k} \\ &\quad \times \Psi_{k,j,F}(t_1, \dots, t_k), \end{aligned} \quad (20)$$

where  $G(t) = \pi_0 t + (1 - \pi_0)F(t)$ . For any  $m_0 \in \{0, \dots, m\}$ ,  $k \geq 0$ ,  $k \leq m$ ,  $j \leq m_0$ ,  $k - j \leq m - m_0$ , we let

$$\begin{aligned} \mathcal{Q}_{m,m_0,F}(\mathbf{t}, k, j) &= \binom{m_0}{j} \binom{m-m_0}{k-j} (t_k)^j (F(t_k))^{k-j} \\ &\quad \times \Psi_{m-k,m_0-j,\bar{F}}(1 - t_m, \dots, 1 - t_{k+1}); \end{aligned} \quad (21)$$

$$\begin{aligned} \tilde{\mathcal{Q}}_{m,m_0,F}(\mathbf{t}, k, j) &= \binom{m_0}{j} \binom{m-m_0}{k-j} (1 - t_{k+1})^{m_0-j} (1 - F(t_{k+1}))^{m-m_0-k+j} \\ &\quad \times \Psi_{k,j,F}(t_1, \dots, t_k), \end{aligned} \quad (22)$$

where  $\bar{F}(t) = 1 - F(1 - t)$ . The following results have been proved by Roquain and Villers (2011a,b).

**Theorem 5.2** (Roquain and Villers (2011)). *Consider any threshold collection  $\mathbf{t}$  and the quantities defined by (19), (20), (21) and (22). Then the following holds:*

(i) *In the  $\text{RM}(m, \pi_0, F)$  model, for any  $\pi_0 \in [0, 1]$ ,  $F \in \mathcal{F}$ ,  $0 \leq k \leq m$ ,  $0 \leq j \leq k$ ,*

$$\mathbb{P}(|R \cap \{1, \dots, m_0\}| = j, |R| = k) = \begin{cases} \mathcal{P}_{m,\pi_0,F}(\mathbf{t}, k, j) & \text{for } R = SU(\mathbf{t}), \\ \tilde{\mathcal{P}}_{m,\pi_0,F}(\mathbf{t}, k, j) & \text{for } R = SD(\mathbf{t}). \end{cases} \quad (23)$$

(ii) *In the  $\text{FM}(m, m_0, F)$  model, for any  $m_0 \in \{0, \dots, m\}$ ,  $F \in \mathcal{F}$ ,  $0 \leq k \leq m$ ,  $0 \vee (k - m + m_0) \leq j \leq m_0 \wedge k$ ,*

$$\mathbb{P}(|R \cap \{1, \dots, m_0\}| = j, |R| = k) = \begin{cases} \mathcal{Q}_{m,m_0,F}(\mathbf{t}, k, j) & \text{for } R = SU(\mathbf{t}), \\ \tilde{\mathcal{Q}}_{m,m_0,F}(\mathbf{t}, k, j) & \text{for } R = SD(\mathbf{t}). \end{cases} \quad (24)$$

Classically, any step-up-down procedure can be written as a combination of a step-down and a step-up procedure (Sarkar, 2002):

$$\text{SUD}_\lambda(\mathbf{t}) = \begin{cases} \text{SU}((t_\lambda \wedge t_j)_{1 \leq j \leq m}) & \text{if } |\text{SU}((t_\lambda \wedge t_j)_{1 \leq j \leq m})| < \lambda, \\ \text{SD}((t_\lambda \vee t_j)_{1 \leq j \leq m}) & \text{if } |\text{SD}((t_\lambda \vee t_j)_{1 \leq j \leq m})| \geq \lambda. \end{cases} \quad (25)$$

Moreover, since  $\{|\text{SU}((t_\lambda \wedge t_j)_{1 \leq j \leq m})| < \lambda\} = \{p_{(\lambda)} > t_\lambda\}$  and  $\{|\text{SD}((t_\lambda \vee t_j)_{1 \leq j \leq m})| \geq \lambda\} = \{p_{(\lambda)} \leq t_\lambda\}$  the two cases in (25) form a partition of the probability space. This yields the following explicit FDR computations:

**Corollary 5.3.** *let  $\lambda \in \{1, \dots, m\}$  and consider any threshold collection  $\mathbf{t}$ . Then the following holds:*

(i) *In the model  $\text{RM}(m, \pi_0, F)$ , for any  $\pi_0 \in [0, 1]$ ,  $F \in \mathcal{F}$ , we have*

$$\begin{aligned} \text{FDR}(\text{SUD}_\lambda(\mathbf{t})) &= \sum_{k=1}^{\lambda-1} \sum_{j=0}^k \frac{j}{k} \mathcal{P}_{m, \pi_0, F}(\mathbf{t} \wedge t_\lambda, k, j) \\ &\quad + \sum_{k=\lambda}^m \sum_{j=0}^k \frac{j}{k} \tilde{\mathcal{P}}_{m, \pi_0, F}(\mathbf{t} \vee t_\lambda, k, j). \end{aligned} \quad (26)$$

(ii) *In the model  $\text{FM}(m, m_0, F)$ , for any  $m_0 \in \{0, \dots, m\}$ ,  $F \in \mathcal{F}$ , we have*

$$\begin{aligned} \text{FDR}(\text{SUD}_\lambda(\mathbf{t})) &= \sum_{k=1}^{\lambda-1} \sum_{j=0 \vee (k-m+m_0)}^{m_0 \wedge k} \frac{j}{k} \mathcal{Q}_{m, m_0, F}(\mathbf{t} \wedge t_\lambda, k, j) \\ &\quad + \sum_{k=\lambda}^m \sum_{j=0 \vee (k-m+m_0)}^{m_0 \wedge k} \frac{j}{k} \tilde{\mathcal{Q}}_{m, m_0, F}(\mathbf{t} \vee t_\lambda, k, j). \end{aligned} \quad (27)$$

In the model  $\text{RM}(m, \pi_0, F)$ , it turns out that  $\text{FDR}(\text{SUD}_\lambda(\mathbf{t}))$  has an expression only involving the  $\Psi_k$ s, and not the  $\Psi_{k,j,F}$  (Roquain and Villers, 2011a, Section 5.2). Although it has a somewhat less intuitive form, it is better than (26) from a computational point of view.

## 6 Discussion

Our aim in this paper was to address the question “is the Dirac-uniform distribution an LFC for an intermediate step-up-down procedure (that uses a standard

threshold collection)?" In a nutshell, the answer we found is "no, but almost". We provided a rigorous quantification of what "almost" means, using an alternative approach to the asymptotic results of Gontscharuk (2010) that entails nonasymptotic bounds and explicit convergence rates. In practical situations, evaluating such bounds can allow to determine whether we can consider that the FDR is maximum when the signal strength is maximum.

Returning to equations (5) and (6), an additional question, particularly relevant in practice, is how appropriate it is to base the multiple type I error criterion solely on control of the *expectation* of the random variable FDP. We notice that Theorem 5.2 may also be used to study this issue by computing exactly the point mass function of the FDP under arbitrary configurations for the alternative, cf. Section A in the appendix. Based on this, we investigated to what extent the distribution of the FDP concentrates around its expectation for a simple Gaussian location model with parameter  $\mu$ . Figure 5 was obtained from these exact formulas for  $m = 100$  and varying values of  $\pi_0$  and  $\mu$ . Note that the unrealistically large choice of  $\alpha = 1/2$  has only been used for reasons of readability of the figures; similar plots also obtained when choosing  $\alpha$  smaller (the variance of the FDP actually increases with smaller  $\alpha$ , because this entails a smaller number of rejections). On inspection of these graphs, it becomes apparent that – even though joint independence of the  $p$ -values holds – the distribution of the FDP is not concentrated around the corresponding FDR in the following two situations: (i) The effect size  $\mu$  is close to zero (weak signal) or (ii) the proportion  $\pi_0$  of true null hypotheses is close to 1 (sparse signal). Thus, controlling the FDR alone does not guarantee a small FDP for a specific experiment at hand in these cases. For a well-defined dependency structure induced by exchangeable test statistics, theoretical arguments for  $m$  tending to infinity support the observation that the distribution of the FDP often does not degenerate in the limit, see (Finner et al., 2007; Delattre and Roquain, 2011). For the jointly independent case and in the cases (i) or (ii) above, this phenomenon has not been theoretically studied to the best of our knowledge. The latter can possibly be investigated by extending the asymptotic approach of Neuvial (2008) to the case where  $\mu$  and  $\pi_0$  are allowed to depend on  $m$ .

Taking these considerations into account, control of the false discovery ex-

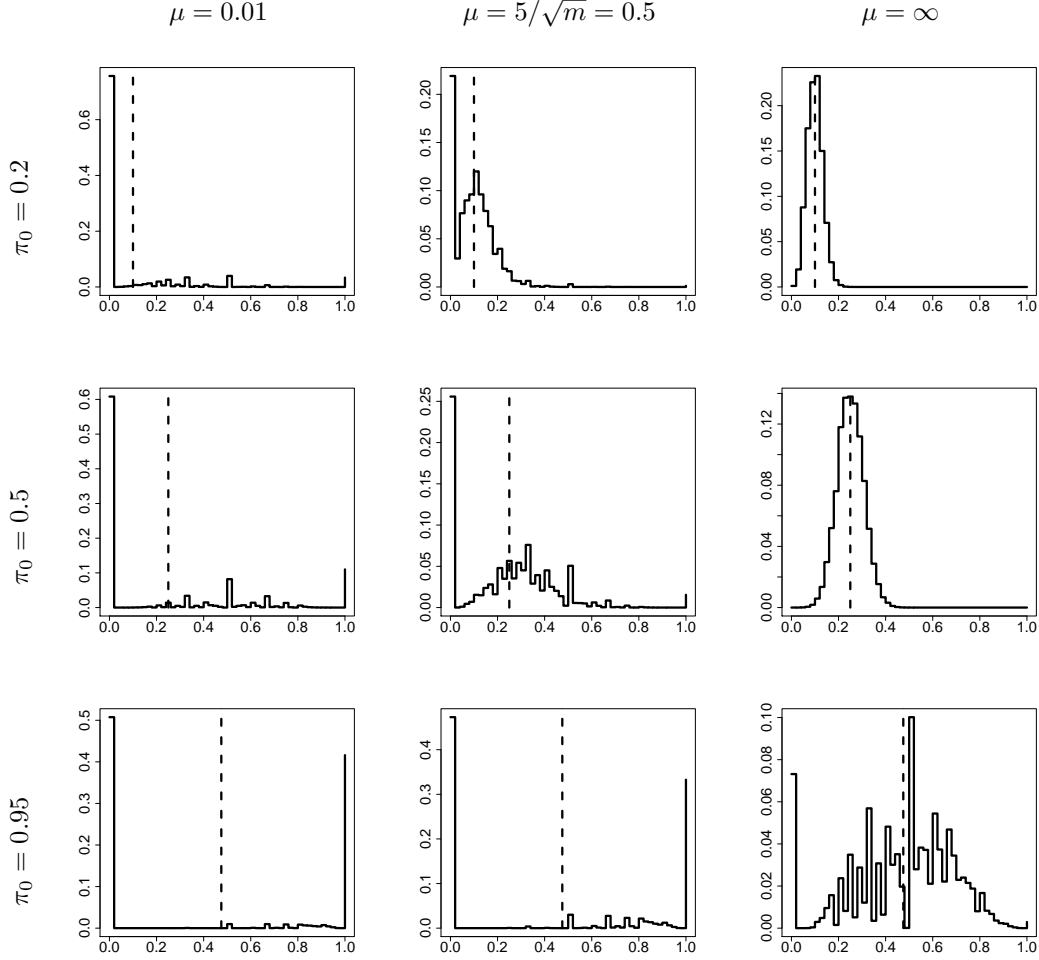


Figure 5: Exact probability  $\mathbb{P}(\text{FDP}(\text{LSU}) \in [i/50, (i+1)/50))$  for  $0 \leq i \leq 50$ . The value of  $\text{FDR}(\text{LSU}) = \pi_0 \alpha$  is displayed by the vertical dashed line. Random mixture model.  $\alpha = 0.5$ ;  $m = 100$ . One sided location Gaussian c.d.f. with parameter  $\mu$ .

ceedance (i.e., of the probability that the FDP exceeds a given threshold) has recently been proposed in the literature (see, e.g. Farcomeni, 2008, for a review). Controlling the false discovery exceedance control again brings forth the question of the corresponding LFC: are Dirac-uniform configurations least favorable for, e.g.,  $\mathbb{P}(\text{FDP}(\text{LSU}) > x)$ ? Some non-reported figures show that this is not the case for any  $x$ . Hence, finding LFCs for the false discovery exceedance stays an open avenue for future research.

## 7 Proofs

### 7.1 Proof of Proposition 5.1

We follow the proof of the regular Steck's recursion (Shorack and Wellner, 1986, p. 366–369). By using the convention  $U_{(0)} = t_0 = 0$  and by considering the smallest  $j$  for which  $U_{(j+1)} > t_{j+1}$ , we can write

$$\begin{aligned} & \mathbb{P}(U_{(k)} \leq t_k) - \mathbb{P}[U_{(1)} \leq t_1, \dots, U_{(k)} \leq t_k] \\ &= \sum_{j=0}^{k-2} \mathbb{P}(\forall i \leq j, U_{(i)} \leq t_i, U_{(j+1)} > t_{j+1}, U_k \leq t_k) \\ &= \sum_{j=0}^{k-2} \sum_{X \subset \{1, \dots, k\}, |X|=j} \mathbb{P}(\forall i \leq j, U_{(i)} \leq t_i, \forall i \notin X, t_{j+1} \leq U_i \leq t_k). \end{aligned}$$

Hence, if  $U_{(i:X)}$  denotes the  $i$ -th smallest member of the set  $\{U_i, i \in X\}$ , we obtain

$$\begin{aligned} & \mathbb{P}(U_{(k)} \leq t_k) - \mathbb{P}[U_{(1)} \leq t_1, \dots, U_{(k)} \leq t_k] \\ &= \sum_{j=0}^{k-2} \sum_{X \subset \{1, \dots, k\}, |X|=j} \mathbb{P}(\forall i \leq j, U_{(i:X)} \leq t_i) \mathbb{P}(\forall i \notin X, t_{j+1} \leq U_i \leq t_k) \\ &= \sum_{j=0}^{k-2} \sum_{j_0=0}^j \sum_{X \subset \{1, \dots, k\}, |X|=j} \mathbf{1}\{|X \cap \{1, \dots, k_0\}| = j_0\} \Psi_{j,j_0,F}(t_1, \dots, t_j) \\ & \quad \times \mathbb{P}(\forall i \notin X, t_{j+1} \leq U_i \leq t_k). \end{aligned}$$

### 7.2 Proof of Lemma 4.3

In this proof, we denote  $\hat{\mathbb{G}}'_m = (\hat{\mathbb{G}}_m + m^{-1}) \wedge 1$  for short. Let us first note that for any  $k \in \{1, \dots, m\}$ , “ $p_{(k)} \leq t_k$ ” is equivalent to “ $\hat{\mathbb{G}}_m(\rho(k/m)) \geq k/m$ ”. We now distinguish two cases:

- Step-up case: assume  $p_{(\lambda)} > t_\lambda$ , that is,  $\hat{\mathbb{G}}_m(\rho(\lambda/m)) < \lambda/m$ . Let us prove that  $\hat{k}/m = \mathcal{U}(\lambda/m, \hat{\mathbb{G}}_m)$ . Since  $\mathcal{U}(\ell, G)$  is a fixed point of the function  $G \circ \rho$ , we have  $\mathcal{U}(\ell, G) \in \{0, 1/m, \dots, m/m\}$ . Hence,

$$\begin{aligned} \hat{k}/m &= \max\{k/m \in \{0, \dots, \lambda/m\} : \hat{\mathbb{G}}_m(\rho(k/m)) \geq k/m\} \\ &= \max\{u \in [0, \lambda/m] : \hat{\mathbb{G}}_m(\rho(u)) \geq u\}, \end{aligned}$$

and we can conclude.

- Step-down case: assume  $p_{(\lambda)} \leq t_\lambda$ , that is,  $\hat{\mathbb{G}}_m(\rho(\lambda/m)) \geq \lambda/m$ . First assume that  $\hat{k} < m$  and prove that  $(\hat{k} + 1)/m = \mathcal{U}(\lambda/m, \hat{\mathbb{G}}'_m)$ . On the one hand,

$$\begin{aligned} (\hat{k} + 1)/m &= \min\{k/m \in \{(\lambda + 1)/m, \dots, m/m\} : \hat{\mathbb{G}}_m(\rho(k/m)) < k/m\} \\ &= \min\{k/m \in \{\lambda/m, \dots, m/m\} : \hat{\mathbb{G}}_m(\rho(k/m)) < k/m\} \\ &= \min\{k/m \in \{\lambda/m, \dots, m/m\} : \hat{\mathbb{G}}'_m(\rho(k/m)) \leq k/m\}, \quad (28) \end{aligned}$$

because  $m\hat{\mathbb{G}}_m(\rho(k/m))$  is an integer. On the other hand, since  $\hat{\mathbb{G}}'_m(\rho(\lambda/m)) \geq \lambda/m$ , we have

$$\begin{aligned} \mathcal{U}(\lambda/m, \hat{\mathbb{G}}'_m) &= \min\left\{u \in [\lambda/m, 1] : \hat{\mathbb{G}}'_m(\rho(u)) \leq u\right\} \\ &= \min\left\{u \in \{\lambda/m, \dots, m/m\} : \hat{\mathbb{G}}'_m(\rho(u)) \leq u\right\}, \quad (29) \end{aligned}$$

because  $\mathcal{U}(\lambda/m, \hat{\mathbb{G}}'_m) \in \{0, 1/m, \dots, m/m\}$ . Combining (28) and (29) yields the result. Second, in the case where  $\hat{k} = m$ , then for any  $k/m \in \{\lambda/m, \dots, m/m\}$ , we have  $\hat{\mathbb{G}}_m(\rho(k/m)) \geq k/m$ . Hence, for all  $k/m \in \{\lambda/m, \dots, (m - 1)/m\}$ ,  $\hat{\mathbb{G}}'_m(\rho(k/m)) > k/m$  which entails  $\mathcal{U}(\lambda/m, \hat{\mathbb{G}}'_m) = 1$ . Hence,  $\hat{k}/m = \mathcal{U}(\lambda/m, \hat{\mathbb{G}}'_m)$  in that case. Finally, the inequality  $\hat{k}/m \leq \mathcal{U}(\lambda/m, \hat{\mathbb{G}}'_m)$  always holds.

### 7.3 Proof of Theorem 4.4

Let us first prove the result in  $\text{FM}(m, m_0, F)$  model. We recall  $\hat{\mathbb{G}}_m(x) := m^{-1} \sum_{i=1}^m \mathbf{1}\{p_i \leq x\}$ ,  $G_\zeta^{DU}(x) := (1 - \zeta) + \zeta x$  and we put  $\hat{u} := \hat{k}/m$ , where  $\hat{k}$  is defined by (2). We can easily check from the definition (9) that if  $G, G'$  are two nondecreasing functions such that  $\forall x \in [0, 1], G(x) \geq G'(x)$ , then  $\mathcal{U}(\lambda, G) \geq \mathcal{U}(\lambda, G')$ . Based on the bound (10), we deduce that  $\forall \delta \in (0, 1)$ ,

$$\left\{ \sup_{x \in [0, 1]} |\hat{\mathbb{G}}_m(x) - G_\zeta^{DU}(x)| \leq \delta - \frac{1}{m} \right\} \subset \{u_\delta^- \leq \hat{u} \leq u_\delta^+\}.$$



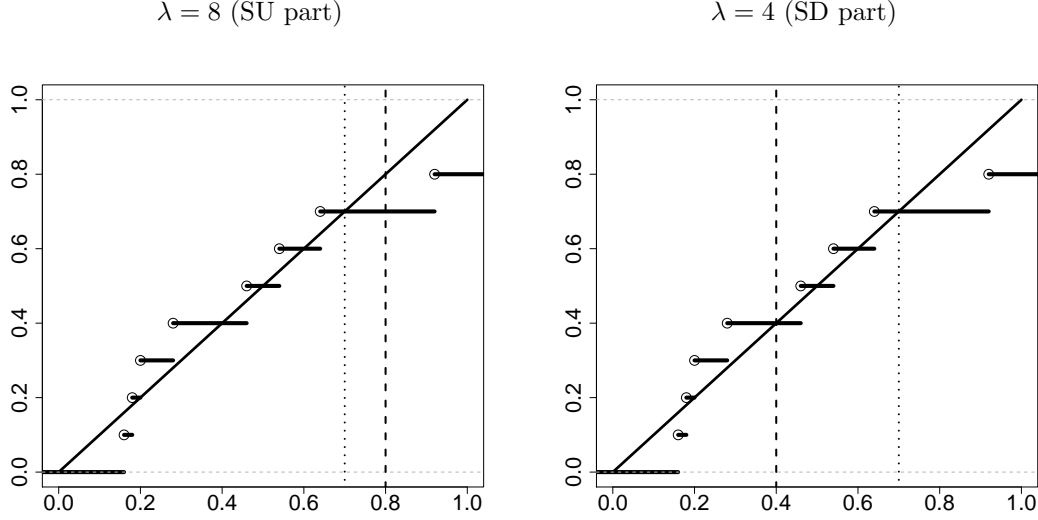


Figure 6: Plot of  $\mathbb{G}_m(\rho(\cdot))$  for  $\rho(x) = 0.5x$  and  $m = 10$   $p$ -values. The order  $\lambda$  of the SUD procedure is displayed by the dashed line while the value of  $\hat{k}/m$  is displayed by the dotted line. Left: the SUD procedure of order  $\lambda = 8$  is such that  $\hat{k}/m = \mathcal{U}(\lambda/m, \hat{\mathbb{G}}_m) = 0.7$ . Right: the SUD procedure of order  $\lambda = 4$  satisfies  $\hat{k}/m = 0.7$  but  $\mathcal{U}(\lambda/m, \hat{\mathbb{G}}_m) = 0.4$ .

As a consequence, in the  $\text{DU}(m, k)$  model,  $k \in \{m_0 - 1, m_0\}$ , by using  $\hat{\mathbb{G}}_m(x) - G_\zeta^{\text{DU}}(x) = (k/m - \zeta)(1 - \hat{\mathbb{G}}_k(x)) - \zeta(\hat{\mathbb{G}}_k(x) - x)$ , we obtain

$$\Omega_\delta(k) := \left\{ \sup_{x \in [0,1]} |\hat{\mathbb{G}}_k(x) - x| \leq \zeta^{-1}(\delta - \nu - \frac{1}{m}) \right\} \subset \{u_\delta^- \leq \hat{u} \leq u_\delta^+\}. \quad (30)$$

Remember that  $p_1, \dots, p_{m_0}$  correspond to true nulls, hence, when  $k \in \{m_0 - 1, m_0\}$ ,  $\hat{\mathbb{G}}_k$  involves only variables which are i.i.d. uniform. As a consequence, by using the DKW inequality with Massart's (1990) optimal constant, we have in the  $\text{DU}(m, k)$  model and for  $k \in \{m_0 - 1, m_0\}$ ,

$$\begin{aligned} \mathbb{P}_{\text{DU}(m,k)}[(\Omega_\delta(k))^c] &\leq 2 \exp \left\{ -2k \left( \delta - \nu - \frac{1}{m} \right)_+^2 / \zeta^2 \right\} \\ &\leq 2 \exp \left\{ -2m \left( \delta - \nu - \frac{1}{m} \right)_+^2 (1 - \nu/\zeta)_+ \right\}, \end{aligned} \quad (31)$$

because  $k/m \geq \zeta - \nu$  and  $\zeta \leq 1$ .

### 7.3.1 Upper bound

Let  $q(x) = \rho(x)/x$  when  $x \in (0, 1]$  and  $q(0) = \lim_{x \rightarrow 0+} \rho(x)/x$  (the limit exists in  $\mathbb{R}$  because  $x \in (0, 1] \rightarrow \rho(x)/x$  is non-decreasing). Applying Theorem 4.3 of Finner et al. (2009), we have

$$\begin{aligned} \text{FDR}(\text{SUD}_\lambda(\mathbf{t}), m_0, F) &\leq \frac{m_0}{m} \mathbb{E}_{\text{DU}(m, m_0-1)}[q(\hat{u})] \\ &\leq \frac{m_0}{m} \frac{\rho(u_\delta^+)}{u_\delta^+} + \frac{m_0}{m} \mathbb{P}_{\text{DU}(m, m_0-1)}[(\Omega_\delta(m_0-1))^c], \end{aligned} \quad (32)$$

because  $q$  is non-decreasing,  $u_\delta^+$  is positive and by (30). Next, by (31), we obtain the following upper-bound:

$$\text{FDR}(\text{SUD}_\lambda(\mathbf{t}), m_0, F) \leq \frac{m_0}{m} \frac{\rho(u_\delta^+)}{u_\delta^+} + \frac{m_0}{m} 2 \exp \left\{ -2m \left( \delta - \nu - \frac{1}{m} \right)_+^2 (1 - \nu/\zeta)_+ \right\}. \quad (33)$$

### 7.3.2 Lower bound

In the model  $\text{DU}(m, m_0)$  with  $m_0 < m$ , we have  $\hat{u} > 0$  a.s. and thus

$$\begin{aligned} \text{FDR}(\text{SUD}_\lambda(\mathbf{t}), m_0, F \equiv 1) &= \frac{m_0}{m} \mathbb{E}_{\text{DU}(m, m_0)} \left( \frac{\hat{\mathbb{G}}_{m_0}(\rho(\hat{u}))}{\hat{u}} \right) \\ &\geq \frac{m_0}{m} \mathbb{E}_{\text{DU}(m, m_0)} \left( \frac{\hat{\mathbb{G}}_{m_0}(\rho(\hat{u}))}{\hat{u}} \mathbf{1}_{\{\Omega_\delta(m_0)\}} \right) \\ &\geq \frac{m_0}{m} \mathbb{E}_{\text{DU}(m, m_0)} \left( \frac{\hat{\mathbb{G}}_{m_0}(\rho(u_\delta^-))}{u_\delta^+} \mathbf{1}_{\{\Omega_\delta(m_0)\}} \right) \\ &\geq \frac{m_0}{m} \frac{\rho(u_\delta^-)}{u_\delta^+} - \frac{m_0}{m} \frac{1}{1-\zeta} \mathbb{P}_{\text{DU}(m, m_0)}[(\Omega_\delta(m_0))^c], \end{aligned}$$

by (30) and because  $u_\delta^+ \geq 1 - \zeta$ . From (31), we obtain the lower-bound

$$\text{FDR}(\text{SUD}_\lambda(\mathbf{t}), m_0, F \equiv 1) \geq \frac{m_0}{m} \frac{\rho(u_\delta^-)}{u_\delta^+} - \frac{m_0}{m} \frac{1}{1-\zeta} 2 \exp \left\{ -2m \left( \delta - \nu - \frac{1}{m} \right)_+^2 (1 - \nu/\zeta)_+ \right\}. \quad (34)$$

Finally, (33) and (34) yield the result.

### 7.3.3 Proof for random mixture model

In the  $\text{RM}(m, \pi_0, F)$  model with  $\pi_0 = \zeta$ , the distribution of  $m_0$  is binomial with parameters  $(m, \zeta)$ . In particular,  $\nu$  is random. However, we can write

$$\begin{aligned} \mathbb{E}[\text{FDP}(\text{SUD}_\lambda(\mathbf{t}), m_0)] &\leq \mathbb{E}[\text{FDP}(\text{SUD}_\lambda(\mathbf{t}), m_0) \mathbf{1}\{\nu \leq \gamma\}] \\ &\quad + 2 \mathbb{P}(|m_0/m - \pi_0| > \gamma - 1/m). \end{aligned} \quad (35)$$

Additionally, using Hoeffding's (1963) inequality, we can write

$$\mathbb{P}(|m_0/m - \pi_0| > \gamma - 1/m) \leq 2e^{-2m(\gamma - 1/m)_+^2}. \quad (36)$$

Combining (35) and (36) with (12) finishes the proof.

### 7.4 Proof of Corollary 4.5

First consider the  $\text{FM}(m, m_0, F)$  model with  $m_0 < m$  (the case  $m_0 = m$  is trivial). Let  $\nu_m = \max_{k \in \{m_0 - 1, m_0\}} \{|k/m - \zeta_m|\}$  and consider  $\delta_m \in (\nu_m, 1)$  that satisfies for large  $m$ ,

$$2(1 - \nu_m/\zeta_m)(\delta_m - \nu_m - 1/m)^2 = (\log m)/m, \quad (37)$$

so that  $e^{-2m(\delta_m - \nu_m - 1/m)_+^2(1 - \nu_m/\zeta_m)_+} = 1/m$  for large  $m$ . Since  $\nu_m \leq 2/m$  by assumption, we have  $\delta_m \propto \sqrt{(\log m)/m}$ . From Theorem 4.4, it is sufficient to prove

$$\frac{\rho(u_{\delta_m}^+) - \rho(u_{\delta_m}^-)}{u_{\delta_m}^+} = O(\delta_m/(1 - \zeta_m)).$$

From Section 4.2, this holds for the linear critical value function. This also holds for the AORC as soon as  $\lambda_m/m < v_{\delta_m}$ , which is the case for large  $m$  by assumption.

The proof in the  $\text{RM}(m, \pi_0, F)$  model is similar by taking additionally  $\gamma_m \propto \sqrt{(\log m)/m}$  such that  $2e^{-2m(\gamma_m - 1/m)_+^2} = 1/m$ .

## Acknowledgements

This work was supported by the French Agence Nationale de la Recherche (ANR grant references: ANR-09-JCJC-0027-01, ANR-PARCIMONIE, ANR-09-JCJC-0101-01) and the French ministry of foreign and european affairs (EGIDE - PROCOPE project number 21887 NJ).

## A Formulas for FDP distribution

From Theorem 5.2 and (25), we can compute the exact c.d.f. of the FDP of any SUD procedure in the following way, for each fixed number  $m \geq 2$  of hypotheses.

**Corollary A.1.** *Let  $\lambda \in \{1, \dots, m\}$  and consider any threshold collection  $\mathbf{t}$ . Fix an arbitrary  $x \in (0, 1)$ . Then the following holds:*

(i) *in the model  $\text{RM}(m, \pi_0, F)$ , for any  $\pi_0 \in [0, 1]$ ,  $F \in \mathcal{F}$ , we have*

$$\begin{aligned} \mathbb{P}(\text{FDP}(\text{SUD}_\lambda(\mathbf{t})) \leq x) &= \sum_{k=0}^{\lambda-1} \sum_{j=0}^{\lfloor xk \rfloor} \mathcal{P}_{m, \pi_0, F}((t_\lambda \wedge t_j)_{1 \leq j \leq m}, k, j) \\ &\quad + \sum_{k=\lambda}^m \sum_{j=0}^{\lfloor xk \rfloor} \tilde{\mathcal{P}}_{m, \pi_0, F}((t_\lambda \vee t_j)_{1 \leq j \leq m}, k, j); \end{aligned} \quad (38)$$

(ii) *in the model  $\text{FM}(m, m_0, F)$ , for any  $m_0 \in \{0, \dots, m\}$ ,  $F \in \mathcal{F}$ , we have*

$$\begin{aligned} \mathbb{P}(\text{FDP}(\text{SUD}_\lambda(\mathbf{t})) \leq x) &= \sum_{k=0}^{\lambda-1} \sum_{j=0 \vee (k-m+m_0)}^{m_0 \wedge \lfloor xk \rfloor} \mathcal{Q}_{m, m_0, F}((t_\lambda \wedge t_j)_{1 \leq j \leq m}, k, j) \\ &\quad + \sum_{k=\lambda}^m \sum_{j=0 \vee (k-m+m_0)}^{m_0 \wedge \lfloor xk \rfloor} \tilde{\mathcal{Q}}_{m, m_0, F}((t_\lambda \vee t_j)_{1 \leq j \leq m}, k, j). \end{aligned} \quad (39)$$

## B FDR(SD) can exceed FDR(SU) in an extreme configuration

**Lemma B.1.** *Consider the  $\text{FM}(m, m_0, F)$  model with  $F(x) = \mathbf{1}\{x \geq 1\}$  (i.e., all the  $p$ -values under the alternative are constantly equal to 1). Consider the threshold collection  $\mathbf{t}$  defined by  $t_k = t_0$ ,  $1 \leq k \leq m-1$  and  $t_m = 1$ , for some  $t_0 \in (0, 1)$ . Then we have for any  $\lambda \in \{1, \dots, m-1\}$ ,*

$$\text{FDR}(\text{SUD}_\lambda(\mathbf{t})) = 1 - (1 - t_0)^{m_0};$$

$$\text{FDR}(\text{SU}(\mathbf{t})) = m_0/m.$$

*In particular,  $\text{FDR}(\text{SD}(\mathbf{t})) > \text{FDR}(\text{SU}(\mathbf{t}))$  for  $t_0 > 1 - (1 - m_0/m)^{1/m_0}$ .*

The proof is straightforward and is left to the reader. As an illustration, for  $m = 10$  and  $m_0 = 7$ ,  $1 - (1 - m_0/m)^{1/m_0} \simeq 0.158$ .

## C DU is an LFC for the $k$ -FWER

We state here for the sake of completeness a straightforward generalization of Lemma 1 of Finner and Gontscharuk (2009) (see also Lemma 2.2 of Gontscharuk, 2010) concerning the LFCs of multiple testing procedures under a class of type I criteria containing in particular the  $k$ -FWER (but not the FDR, as pointed out in the introduction). This result should be considered as already known by experts in the field, although we failed to locate a precise reference for it. The setting considered assumes independence of  $p$ -values corresponding to true nulls, but is more general than the fixed mixture model, since the  $p$ -values corresponding to true null hypotheses are only assumed to be stochastically larger than a uniform variable on  $[0,1]$ ; also, the  $p$ -values corresponding to alternatives are not assumed to be identically distributed nor independent.

**Lemma C.1.** *Let  $m \geq 1$  and  $m_0 \in \{0, \dots, m\}$  be fixed. Let  $\mathbf{p} = (p_1, \dots, p_m)$  be a family  $p$ -values with distribution by  $P$  such that  $(p_i)_{1 \leq i \leq m_0}$  form an independent family of variables, each stochastically lower bounded by a uniform variable. Assume that  $\delta$  is a multiple testing procedure rejecting all hypotheses having  $p$ -value less than a data-dependent threshold  $t^*(\mathbf{p})$ . Let  $R$  be a type I error criterion taking the form*

$$R(P, \delta) = \mathbb{E}_{\mathbf{p} \sim P}[\phi(V_m(\delta(\mathbf{p})))],$$

where  $V_m$  is defined in (1) and  $\phi$  is a function from  $\mathbb{N}$  to  $\mathbb{R}$ .

Assume the two following conditions are satisfied:

- (i)  $t^*$  is a nonincreasing function of each  $p$ -value;
- (ii)  $\phi$  is nondecreasing.

Then it holds that

$$R(P, \delta) \leq R(\text{DU}(m, m_0), \delta),$$

that is,  $\text{DU}(m, m_0)$  is an LFC for  $\delta$  among the set of distributions satisfying the properties described above.

*Proof.* Using (i) and (ii) together entails that  $\mathbf{p} \mapsto \phi(V_m(\delta(\mathbf{p})))$  is a nonincreasing function of each  $p$ -value. Denote  $\mathbf{p}_0 = (p_1, \dots, p_{m_0}, 0, \dots, 0)$  the  $p$ -value family obtained by replacing  $p_i$  by 0 for  $i > m_0$ , and  $P_0$  the distribution of  $\mathbf{p}_0$  when  $\mathbf{p}$

has distribution  $P$ . Obviously we have

$$\mathbb{E}_{\mathbf{p} \sim P}[\phi(V_m(\delta(\mathbf{p})))] \leq \mathbb{E}_{\mathbf{p} \sim P}[\phi(V_m(\delta(\mathbf{p}_0)))] = \mathbb{E}_{\mathbf{p} \sim P_0}[\phi(V_m(\delta(\mathbf{p})))]$$

Now applying Lemma A.11 as cited by Gontscharuk (2010), we obtain

$$\mathbb{E}_{\mathbf{p} \sim P_0}[\phi(V_m(\delta(\mathbf{p})))] \leq \mathbb{E}_{\mathbf{p} \sim \text{DU}(m, m_0)}[\phi(V_m(\delta(\mathbf{p})))],$$

and thus the conclusion.  $\square$

A straightforward (though less immediately interpretable) extension of this result to procedures that are not necessarily threshold-based is to replace assumption (i) by (i'):  $\mathbf{p} \mapsto V_m(\delta(\mathbf{p}))$  is a non increasing function of each  $p$ -value.

## References

- Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B*, 57(1): 289–300, 1995. ISSN 0035-9246.
- Y. Benjamini and D. Yekutieli. The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.*, 29(4):1165–1188, 2001. ISSN 0090-5364.
- S. Delattre and E. Roquain. On the false discovery proportion convergence under Gaussian equi-correlation. *Statist. Probab. Lett.*, 81(1):111–115, 2011. ISSN 0167-7152. doi: 10.1016/j.spl.2010.09.025. URL <http://dx.doi.org/10.1016/j.spl.2010.09.025>.
- S. Dudoit and M. J. van der Laan. *Multiple testing procedures with applications to genomics*. Springer Series in Statistics. Springer, New York, 2008. ISBN 978-0-387-49316-9. doi: 10.1007/978-0-387-49317-6. URL <http://dx.doi.org/10.1007/978-0-387-49317-6>.
- A. Farcomeni. A review of modern multiple hypothesis testing, with particular attention to the false discovery proportion. *Stat. Methods Med. Res.*, 17(4): 347–388, 2008. ISSN 0962-2802. doi: 10.1177/0962280206079046. URL <http://dx.doi.org/10.1177/0962280206079046>.

- H. Finner and V. Gontscharuk. Controlling the familywise error rate with plug-in estimator for the proportion of true null hypotheses. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 71(5):1031–1048, 2009. ISSN 1369-7412. doi: 10.1111/j.1467-9868.2009.00719.x. URL <http://dx.doi.org/10.1111/j.1467-9868.2009.00719.x>.
- H. Finner and M. Roters. On the false discovery rate and expected type I errors. *Biom. J.*, 43(8):985–1005, 2001. ISSN 0323-3847.
- H. Finner, T. Dickhaus, and M. Roters. Dependency and false discovery rate: asymptotics. *Ann. Statist.*, 35(4):1432–1455, 2007. ISSN 0090-5364.
- H. Finner, T. Dickhaus, and M. Roters. On the false discovery rate and an asymptotically optimal rejection curve. *Ann. Statist.*, 37(2):596–618, 2009.
- D. H. Glueck, A. Karimpour-Fard, J. Mandel, L. Hunter, and K. E. Muller. Fast computation by block permanents of cumulative distribution functions of order statistics from several populations. *Commun. Stat. Theory Methods*, 37(18):2815–2824, 2008.
- V. Gontscharuk. *Asymptotic and Exact Results on FWER and FDR in Multiple Hypothesis Testing*. PhD thesis, Heinrich-Heine-Universität Düsseldorf, 2010.
- W. Guo and M. B. Rao. On control of the false discovery rate under no assumption of dependency. *J. Statist. Plann. Inference*, 138(10):3176–3188, 2008. ISSN 0378-3758. doi: 10.1016/j.jspi.2008.01.003. URL <http://dx.doi.org/10.1016/j.jspi.2008.01.003>.
- W. Hoeffding. Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.*, 58:13–30, 1963. ISSN 0162-1459.
- E. L. Lehmann and J. P. Romano. Generalizations of the familywise error rate. *Ann. Statist.*, 33:1138–1154, 2005.
- P. Massart. The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *Ann. Probab.*, 18(3):1269–1283, 1990. ISSN 0091-1798.
- C. J. Miller, C. Genovese, R. C. Nichol, L. Wasserman, A. Connolly, D. Reichart, A. Hopkins, J. Schneider, and A. Moore. Controlling the false-discovery rate

- in astrophysical data analysis. *The Astronomical Journal*, 122(6):3492–3505, 2001. URL <http://stacks.iop.org/1538-3881/122/i=6/a=3492>.
- P. Neuvial. Asymptotic properties of false discovery rate controlling procedures under independence. *Electron. J. Stat.*, 2:1065–1110, 2008. ISSN 1935-7524. doi: 10.1214/08-EJS207. URL <http://dx.doi.org/10.1214/08-EJS207>.
- D. Pantazis, T. E. Nichols, S. Baillet, and R. M. Leahy. A comparison of random field theory and permutation methods for statistical analysis of meg data. *NeuroImage*, 25:383–394, 2005.
- J. P. Romano and M. Wolf. Control of generalized error rates in multiple testing. *Ann. Statist.*, 35(4):1378–1408, 2007. ISSN 0090-5364. doi: 10.1214/009053606000001622. URL <http://dx.doi.org/10.1214/009053606000001622>.
- E. Roquain and F. Villers. Exact calculations for false discovery proportion with application to least favorable configurations. *Ann. Statist.*, 39(1):584–612, 2011a. ISSN 0090-5364. doi: 10.1214/10-AOS847. URL <http://dx.doi.org/10.1214/10-AOS847>.
- E. Roquain and F. Villers. Supplement to “Exact calculations for false discovery proportion with application to least favorable configurations”. *Ann. Statist.*, 2011b. Available online.
- S. K. Sarkar. Some results on false discovery rate in stepwise multiple testing procedures. *Ann. Statist.*, 30(1):239–257, 2002. ISSN 0090-5364.
- G. R. Shorack and J. A. Wellner. *Empirical processes with applications to statistics*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986. ISBN 0-471-86725-X.
- R. J. Simes. An improved Bonferroni procedure for multiple tests of significance. *Biometrika*, 73(3):751–754, 1986. ISSN 0006-3444.
- P. N. Somerville and C. Hemmelmann. Step-up and step-down procedures controlling the number and proportion of false positives. *Comput. Statist. Data*



- Anal.*, 52(3):1323–1334, 2008. ISSN 0167-9473. doi: 10.1016/j.csda.2007.07.016. URL <http://dx.doi.org/10.1016/j.csda.2007.07.016>.
- A. C. Tamhane, W. Liu, and C. W. Dunnett. A generalized step-up-down multiple test procedure. *Canad. J. Statist.*, 26(2):353–363, 1998. ISSN 0319-5724. doi: 10.2307/3315516. URL <http://dx.doi.org/10.2307/3315516>.
- A. Zeisel, O. Zuk, and E. Domany. FDR control with adaptive procedures and FDR monotonicity. *Ann. Appl. Stat.*, 2011. (To appear).

Institut für Mathematik, Universität Potsdam, Germany

E-mail: [gilles.blanchard@math.uni-potsdam.de](mailto:gilles.blanchard@math.uni-potsdam.de)

Institut für Mathematik, Humboldt-Universität, Germany

E-mail: [dickhaus@math.hu-berlin.de](mailto:dickhaus@math.hu-berlin.de)

LPMA, UPMC, Université Paris 6, France

E-mail: [etienne.roquain@upmc.fr](mailto:etienne.roquain@upmc.fr)

LPMA, UPMC, Université Paris 6, France

E-mail: [fanny.villers@upmc.fr](mailto:fanny.villers@upmc.fr)